

New Generalizations of the Bethe Approximation via Asymptotic Expansion

Ryuhei Mori and Toshiyuki Tanaka *

Abstract

The Bethe approximation, discovered in statistical physics, gives an efficient algorithm called belief propagation (BP) for approximating a partition function. BP empirically gives an accurate approximation for many problems, e.g., low-density parity-check codes, compressed sensing, etc. Recently, Vontobel gives a novel characterization of the Bethe approximation using graph cover. In this paper, a new approximation based on the Bethe approximation is proposed. The new approximation is derived from Vontobel's characterization using graph cover, and expressed by using the edge zeta function, which is related with the Hessian of the Bethe free energy as shown by Watanabe and Fukumizu. On some conditions, it is proved that the new approximation is asymptotically better than the Bethe approximation.

Keywords

Bethe approximation, belief propagation, partition function, graphical model, edge zeta function, asymptotic expansion.

I. INTRODUCTION

Calculation of partition functions is one of the central problems in statistical physics. The Bethe approximation is an empirically successful efficient approximation for many problems [1]. A few results about asymptotic exactness of the Bethe approximation are recently known. Recently, Vontobel shows a novel characterization of the Bethe approximation using graph cover [2]. In this paper, a series of generalizations of the Bethe approximation is shown from Vontobel's characterization of the Bethe free energy via asymptotic expansion.

II. FACTOR GRAPH AND THE BETHE FREE ENERGY

Let \mathcal{X} be a finite set. A factor graph is a bipartite graph consisting of variable nodes and factor nodes. A factor graph including N variable nodes defines a probability measure on \mathcal{X}^N . The set of variable nodes, the set of factor nodes and the set of edges are denoted by V , F and $E \subseteq V \times F$, respectively. The neighborhoods of a factor node $a \in F$ and a variable node $i \in V$ are denoted by $\partial a \subseteq V$ and $\partial i \subseteq F$, respectively. The degrees of variable node i and factor node a are denoted by d_i and d_a , respectively. For each factor node a , there exists a corresponding non-negative function $f_a : \mathcal{X}^{\partial a} \rightarrow \mathbb{R}_{\geq 0}$. The probability mass function defined by the factor graph is

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{a \in F} f_a(\mathbf{x}_{\partial a}), \quad Z := \sum_{\mathbf{x} \in \mathcal{X}^N} \prod_{a \in F} f_a(\mathbf{x}_{\partial a}).$$

The constant Z for the normalization is called the partition function. The calculation of Z is generally a #P-complete

problem. Approximating Z is one of the central problems of statistical physics and information theory [1].

In the following, the Bethe free energy is defined using the variational method [3]. Let $p(\mathbf{x})$ be a probability mass function defined by a factor graph. From the following equation for any probability measure $q(\mathbf{x})$ on \mathcal{X}^N ,

$$\begin{aligned} D(q||p) &:= \sum_{\mathbf{x} \in \mathcal{X}^N} q(\mathbf{x}) \log \frac{q(\mathbf{x})}{p(\mathbf{x})} \\ &= \log Z - \sum_{a \in F} \sum_{\mathbf{x} \in \mathcal{X}^N} q(\mathbf{x}) \log f_a(\mathbf{x}_{\partial a}) + \sum_{\mathbf{x} \in \mathcal{X}^N} q(\mathbf{x}) \log q(\mathbf{x}) \\ &=: \log Z + \mathcal{U}_{\text{Gibbs}}(q) - \mathcal{H}_{\text{Gibbs}}(q) =: \log Z + F_{\text{Gibbs}}(q) \end{aligned}$$

one obtains the equation $-\log Z = F_{\text{Gibbs}}(p) = \min_{q \in \mathcal{P}(\mathcal{X}^N)} F_{\text{Gibbs}}(q)$ where $\mathcal{P}(\mathcal{X}^N)$ denotes the set of probability measures on \mathcal{X}^N . The Bethe free energy, defined in the following, is an approximation of the Gibbs free energy, which is $F_{\text{Gibbs}}(q)$ in the above equation.

Definition 1 (Bethe free energy). For $\{b_i \in \mathcal{P}(\mathcal{X})\}_{i \in V}$ and $\{b_a \in \mathcal{P}(\mathcal{X}^{\partial a})\}_{a \in F}$ satisfying the reducibility

$$\sum_{\mathbf{x}_{\partial a} \in \mathcal{X}^{\partial a}, \mathbf{x}_i = z} b_a(\mathbf{x}_{\partial a}) = b_i(z), \quad \forall (i, a) \in E \quad (1)$$

the Bethe free energy is defined as

$$F_{\text{Bethe}}(\{b_i\}, \{b_a\}) := \mathcal{U}_{\text{Bethe}}(\{b_i\}, \{b_a\}) - \mathcal{H}_{\text{Bethe}}(\{b_i\}, \{b_a\})$$

where

$$\begin{aligned} \mathcal{U}_{\text{Bethe}}(\{b_i\}, \{b_a\}) &:= - \sum_{a \in F} \sum_{\mathbf{x}_{\partial a} \in \mathcal{X}^{\partial a}} b_a(\mathbf{x}_{\partial a}) \log f_a(\mathbf{x}_{\partial a}) \\ \mathcal{H}_{\text{Bethe}}(\{b_i\}, \{b_a\}) &:= - \sum_{a \in F} \sum_{\mathbf{x}_{\partial a} \in \mathcal{X}^{\partial a}} b_a(\mathbf{x}_{\partial a}) \log b_a(\mathbf{x}_{\partial a}) \\ &\quad + \sum_{i \in V} (d_i - 1) \sum_{x_i \in \mathcal{X}} b_i(x_i) \log b_i(x_i). \quad \square \end{aligned}$$

Since the Bethe free energy is considered as an approximation of the Gibbs free energy, the minimum of the Bethe free energy is regarded as an approximation of $-\log Z$. The stationary condition of the Bethe free energy is

$$\begin{aligned} b_a(\mathbf{x}) &= \frac{1}{Z_a(\{m_{i \rightarrow a}\}_{i \in \partial a})} f_a(\mathbf{x}) \prod_{i \in \partial a} m_{i \rightarrow a}(x_i) \\ b_i(x) &= \frac{1}{Z_i(\{m_{a \rightarrow i}\}_{a \in \partial i})} \prod_{a \in \partial i} m_{a \rightarrow i}(x) \\ &= \frac{1}{Z_{i,a}(m_{a \rightarrow i}, m_{i \rightarrow a})} m_{a \rightarrow i}(x) m_{i \rightarrow a}(x), \quad \forall a \in \partial i. \end{aligned} \quad (2)$$

where $\{(m_{a \rightarrow i}, m_{i \rightarrow a}) \in \mathcal{P}(\mathcal{X})^2\}_{(i,a) \in E}$ satisfies the belief propagation (BP) equations

$$\begin{aligned} m_{i \rightarrow a}(x) &\propto \prod_{b \in \partial i \setminus a} m_{b \rightarrow i}(x) \\ m_{a \rightarrow i}(x) &\propto \sum_{\mathbf{x} \in \mathcal{X}^{\partial a}, \mathbf{x}_i = x} f_a(\mathbf{x}) \prod_{j \in \partial a \setminus i} m_{j \rightarrow a}(x_j) \end{aligned} \quad (3)$$

* Graduate School of Informatics, Kyoto University, Kyoto, 606-8501, Japan (email: rmori@sys.i.kyoto-u.ac.jp, tt@i.kyoto-u.ac.jp). The work of R. Mori was supported by the Grant-in-Aid for Scientific Research for JSPS Fellows (22-5936).

and where $Z_a(\{m_{i \rightarrow a}\}_{i \in \partial a})$, $Z_i(\{m_{a \rightarrow i}\}_{a \in \partial i})$ and $Z_{i,a}(m_{a \rightarrow i}, m_{i \rightarrow a})$ are the constants for the normalizations

$$\begin{aligned} Z_a(\{m_{i \rightarrow a}\}_{i \in \partial a}) &:= \sum_{\mathbf{x} \in \mathcal{X}^{d_a}} f_a(\mathbf{x}) \prod_{i \in \partial a} m_{i \rightarrow a}(x_i) \\ Z_i(\{m_{a \rightarrow i}\}_{a \in \partial i}) &:= \sum_{\mathbf{x} \in \mathcal{X}} \prod_{a \in \partial i} m_{a \rightarrow i}(x) \\ Z_{i,a}(m_{a \rightarrow i}, m_{i \rightarrow a}) &:= \sum_{x \in \mathcal{X}} m_{a \rightarrow i}(x) m_{i \rightarrow a}(x). \end{aligned}$$

By substituting (2) to the definition of the Bethe free energy and using (3), one obtains the representation

$$\begin{aligned} F_{\text{Bethe}}(\{m_{i \rightarrow a}\}, \{m_{a \rightarrow i}\}) &= - \sum_{i \in V} \log Z_i(\{m_{a \rightarrow i}\}_{a \in \partial i}) \\ &- \sum_{a \in F} \log Z_a(\{m_{i \rightarrow a}\}_{i \in \partial a}) + \sum_{(i,a) \in E} \log Z_{i,a}(m_{i \rightarrow a}, m_{a \rightarrow i}). \end{aligned} \quad (4)$$

Note that the stationary condition of (4) is also given by the BP equations (3). Since the Bethe free energy is an approximation of $-\log Z$, the Bethe partition function is defined as $Z_{\text{Bethe}}(\{m_{i \rightarrow a}\}, \{m_{a \rightarrow i}\}) := \exp\{-F_{\text{Bethe}}(\{m_{i \rightarrow a}\}, \{m_{a \rightarrow i}\})\}$ i.e.,

$$\begin{aligned} Z_{\text{Bethe}}(\{m_{i \rightarrow a}\}, \{m_{a \rightarrow i}\}) &= \prod_{i \in V} Z_i(\{m_{a \rightarrow i}\}_{a \in \partial i}) \\ &\cdot \prod_{a \in F} Z_a(\{m_{i \rightarrow a}\}_{i \in \partial a}) \prod_{(i,a) \in E} \frac{1}{Z_{i,a}(m_{i \rightarrow a}, m_{a \rightarrow i})} \end{aligned}$$

for $\{(m_{a \rightarrow i}, m_{i \rightarrow a})\}_{(i,a) \in E}$ satisfying (3). We also use the notation $Z_{\text{Bethe}} := \min_{\{m_{i \rightarrow a}\}, \{m_{a \rightarrow i}\}} Z_{\text{Bethe}}(\{m_{i \rightarrow a}\}, \{m_{a \rightarrow i}\})$. We assume in this paper that all variable nodes have the same alphabet \mathcal{X} only for the simplicity. All results in this paper can be generalized to the case that each variable node $i \in V$ has distinct alphabet \mathcal{X}_i .

III. EXPONENTIAL FAMILY

In this section, exponential family of probability distributions is introduced. The exponential family is a class of parametric family of distributions. A family of probability distributions having a parameter $\theta \in \Theta \subseteq \mathbb{R}^d$ for some $d \in \mathbb{N}$ is called a parametric family of distributions. In this paper, the existence of the first and the second derivatives of probability density (mass) function $p(x | \theta)$ with respect to the parameter is assumed.

Definition 2 (Fisher information matrix). For a parametric family of distributions, the Fisher information matrix is defined for each parameter $\theta \in \Theta$ as

$$\mathcal{J}(\theta)_{k,l} := \mathbb{E} \left[\frac{\partial \log p(X | \theta)}{\partial \theta_k} \frac{\partial \log p(X | \theta)}{\partial \theta_l} \right]$$

where $X \sim p(X | \theta)$ and where $B_{k,l}$ denotes the (k, l) -element of a matrix B .

Definition 3 (Exponential family). The exponential family is a parametric family of probability distributions whose probability density (mass) functions can be expressed in the form

$$p(x | \theta) = \exp \left\{ \sum_{k=1}^d t_k(x) \theta_k - \psi(\theta) \right\}$$

using a set of functions $\{t_k : \mathcal{X} \rightarrow \mathbb{R}\}_{k=1,2,\dots,d}$ called a sufficient statistic. Here,

$$\psi(\theta) := \log \sum_{x \in \mathcal{X}} \exp \left\{ \sum_{k=1}^d t_k(x) \theta_k \right\}.$$

The parameter θ of an exponential family is called a natural parameter. The parameterization with $\{\eta_k := \mathbb{E}[t_k(X)]\}_{k=1,\dots,d}$ is also useful, and the parameter is called an expectation parameter.

Example 4 (Multinomial distribution). The family of distributions on a finite set $\mathcal{X} = \{0, 1, \dots, q-1\}$ can be regarded as the exponential family with a sufficient statistic $\{t_x(x') = \delta_{x,x'}\}_{x \in \mathcal{X} \setminus 0}$. In this case, $\eta_x = p(x | \theta)$ for $x \in \mathcal{X} \setminus 0$.

Let $\nabla^2 Q(\theta)$ be the Hessian matrix of Q at θ . Let $\text{Var}[t(X)]$ be the covariance matrix of $\{t_k(X)\}_{k=1,\dots,d}$ with respect to $p(x | \theta)$. Then, it is easy to see that $\text{Var}[t(X)] = \mathcal{J}(\theta)$.

Lemma 5. It holds $\mathcal{J}(\theta) = \mathcal{J}(\eta)^{-1}$ where $\mathcal{J}(\theta)$ and $\mathcal{J}(\eta)$ are the Fisher information matrices with respect to the natural parameter θ and the expectation parameter η . Furthermore, it holds $-\nabla^2 H(\eta) = \mathcal{J}(\eta)$ and $\nabla^2 \psi(\theta) = \mathcal{J}(\theta)$ where $H(\eta) := -\sum_{x \in \mathcal{X}} p(x | \eta) \log p(x | \eta)$ is the Shannon entropy.

IV. CHARACTERIZATION OF THE BETHE FREE ENERGY USING THE LOOP CALCULUS

A. Linear transforms of the factor functions

In this section, a characterization of the Bethe free energy is described which is obtained by Chernyak and Chertkov [4]. Assume that a function $f_a(\mathbf{x}) : \mathcal{X}^{d_a} \rightarrow \mathbb{R}_{\geq 0}$ is represented by using $\hat{f}_a : \mathcal{Y}^{d_a} \rightarrow \mathbb{R}_{\geq 0}$ and $\phi_{i,a} : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_{\geq 0}$ as

$$f_a(\mathbf{x}_{\partial a}) = \sum_{\mathbf{y}_{\partial a} \in \mathcal{Y}^{d_a}} \hat{f}_a(\mathbf{y}_{\partial a}) \prod_{i \in \partial a} \phi_{i,a}(x_i, y_i) \quad (5)$$

for some set \mathcal{Y} . Let $\delta(x, z)$ be a function which takes 1 if $x = z$ and takes 0 otherwise. When $\mathcal{Y} = \mathcal{X}$, this representation is obtained by letting

$$\hat{f}_a(\mathbf{y}_{\partial a}) = \sum_{\mathbf{x}_{\partial a} \in \mathcal{X}^{d_a}} f_a(\mathbf{x}_{\partial a}) \prod_{i \in \partial a} \hat{\phi}_{i,a}(y_i, x_i)$$

for $\{(\phi_{i,a}, \hat{\phi}_{i,a})\}_{i \in \partial a}$ satisfying

$$\sum_{y \in \mathcal{X}} \phi_{i,a}(x, y) \hat{\phi}_{i,a}(y, z) = \delta(x, z)$$

or equivalently

$$\sum_{x \in \mathcal{X}} \hat{\phi}_{i,a}(y, x) \phi_{i,a}(x, w) = \delta(y, w). \quad (6)$$

The partition function can be rewritten by using the transform (5) as

$$Z = \sum_{\mathbf{y} \in \mathcal{Y}^E} \prod_{a \in F} \hat{f}_a(\mathbf{y}_{\partial a}) \prod_{i \in V} \left(\sum_{x \in \mathcal{X}} \prod_{a \in \partial i} \phi_{i,a}(x, y_{i,a}) \right). \quad (7)$$

In this representation, the variables $\mathbf{y} \in \mathcal{Y}^E$ are associated with edges of the original factor graph. Both the variable nodes and factor nodes in the original graph can be regarded as factor nodes in the new representation (7). The transform of representation gives a unified way to understand many equations, e.g., the MacWilliams identity [5].

B. Bethe transforms and characterization of Bethe free energy

Let $\mathcal{Y} = \mathcal{X}$. We now consider the following additional conditions on $\{(\phi_{i,a}, \hat{\phi}_{i,a})\}_{(i,a) \in E}$. For each $i \in V$ and $a \in F$,

$$\sum_{x \in \mathcal{X}} \prod_{a \in \partial i} \phi_{i,a}(x, y_a) = 0, \quad \exists! b \in \partial i, y_b \neq 0$$

$$\sum_{x_{\partial a} \in \mathcal{X}^{d_a}} f_a(x_{\partial a}) \prod_{i \in \partial a} \hat{\phi}_{i,a}(y_i, x_i) = 0, \quad \exists! j \in \partial a, y_j \neq 0. \quad (8)$$

On these conditions, if the subset $\{(i,a) \in E \mid y_{i,a} \neq 0\} \subseteq E$ of edges generates degree-one variable nodes or degree-one factor nodes, the term in (7) corresponding to \mathbf{y} is zero. Hence, in (7), we only have to take the sum over $\mathbf{y} \in \mathcal{X}^{|E|}$ satisfying $\{(i,a) \in E \mid y_{i,a} \neq 0\} \in G$ where G is the set of generalized loops defined as $G := \{E' \subseteq E \mid d_o(E') \neq 1, \forall o \in V \cup F\}$. Here, $d_i(E') := |\{(i,a) \in E' \mid a \in \partial i\}|$ and $d_a(E') := |\{(i,a) \in E' \mid i \in \partial a\}|$. The conditions (6) and (8) are equivalent to the condition (6) together with

$$\hat{\phi}_{i,a}(0, x) = \frac{1}{\sum_{x \in \mathcal{X}} \prod_{b \in \partial i} \phi_{i,b}(x, 0)} \prod_{b \in \partial i \setminus a} \phi_{i,b}(x, 0)$$

$$\phi_{i,a}(x, 0) = \frac{1}{\hat{f}_a(\text{all } 0)} \sum_{x_{\partial a} \in \mathcal{X}^{d_a}} f_a(x_{\partial a}) \prod_{j \in \partial a \setminus i} \hat{\phi}_{j,a}(0, x_j). \quad (9)$$

For $\{(m_{a \rightarrow i}, m_{i \rightarrow a})\}_{(i,a) \in E}$ which satisfies the BP equations (3), $(\phi_{i,a}(x, 0), \hat{\phi}_{i,a}(0, x)) = (c_{i,a} m_{a \rightarrow i}(x), \hat{c}_{i,a} m_{i \rightarrow a}(x))$ provides a solution of (9) where $c_{i,a} \hat{c}_{i,a} = 1/Z_{i,a}(m_{i \rightarrow a}, m_{a \rightarrow i})$. In this case, the contribution of all-zero assignment in (7) is the Bethe partition function $Z_{\text{Bethe}}(\{m_{i \rightarrow a}\}, \{m_{a \rightarrow i}\})$.

For the binary case i.e., $\mathcal{X} = \{0, 1\}$, $(\phi_{i,a}(x, 1), \hat{\phi}_{i,a}(1, x)) = ((-1)^{\bar{x}} c_{i,a} m_{i \rightarrow a}(\bar{x}), (-1)^{\bar{x}} \hat{c}_{i,a} m_{a \rightarrow i}(\bar{x}))$ satisfies (6). In this case, one obtains the following lemma by substituting the above values of $\{\phi_{i,a}, \hat{\phi}_{i,a}\}_{(i,a) \in E}$ to (7). Let $\langle \cdot \rangle_p$ be an expectation with respect to a probability mass function p .

Lemma 6 ([6], [7]). *Assume that the alphabet is binary, i.e., $\mathcal{X} = \{0, 1\}$. Let $m_i := \langle X_i \rangle_{b_i} = b_i(1)$. For any stationary point $(\{b_i\}, \{b_a\})$ of the Bethe free energy,*

$$\frac{Z}{Z_{\text{Bethe}}(\{b_i\}, \{b_a\})} = \sum_{E' \subseteq E} \mathcal{Z}(E') \quad (10)$$

where

$$\mathcal{Z}(E') := \prod_{i \in V} \left\langle \left(\frac{X_i - m_i}{\sqrt{\langle (X_i - m_i)^2 \rangle_{b_i}}} \right)^{d_i(E')} \right\rangle_{b_i}$$

$$\cdot \prod_{a \in F} \left\langle \prod_{i \in \partial a, (i,a) \in E'} \frac{X_i - m_i}{\sqrt{\langle (X_i - m_i)^2 \rangle_{b_i}}} \right\rangle_{b_a}.$$

For non-binary cases, the conditions (6) and (8) do not fix $\{(\phi_{i,a}, \hat{\phi}_{i,a})\}_{(i,a) \in E}$ uniquely. In [4], it is suggested to use loop calculus iteratively for each $\mathcal{Z}(E')$. In this paper, on the other hand, we propose an explicit choice of $\{\phi_{i,a}, \hat{\phi}_{i,a}\}$. As shown in Example 4, the family of multinomial distributions can be regarded as an exponential family. Let θ_i and η_i be a natural parameter and an expectation parameter of b_i , respectively.

Then, $\phi_{i,a}(x, y)$ and $\hat{\phi}_{i,a}(x, y)$ for $x \in \mathcal{X}$ and $y \in \mathcal{X} \setminus 0$ are chosen as

$$\frac{\phi_{i,a}(x, y)}{c_{i,a} m_{a \rightarrow i}(x)} = \frac{\partial \log b_i(x)}{\partial \eta_{i,y}}, \quad \frac{\hat{\phi}_{i,a}(x, y)}{\hat{c}_{i,a} m_{i \rightarrow a}(x)} = \frac{\partial \log b_i(x)}{\partial \theta_{i,y}}.$$

The partial derivatives in the first and second equations in the above are those with respect to the coordinate systems $\{\eta_{i,y}\}_{y \in \mathcal{X} \setminus 0}$ and $\{\theta_{i,y}\}_{y \in \mathcal{X} \setminus 0}$, respectively. One can easily confirm that these $\{\phi_{i,a}, \hat{\phi}_{i,a}\}$ satisfy the condition (6) as follows. For $w \in \mathcal{X} \setminus 0$, it holds

$$\sum_{x \in \mathcal{X}} \hat{\phi}_{i,a}(0, x) \phi_{i,a}(x, w) = \sum_{x \in \mathcal{X}} b_i(x) \frac{\partial \log b_i(x)}{\partial \theta_{i,w}} = 0.$$

Similarly, $\sum_{x \in \mathcal{X}} \hat{\phi}_{i,a}(y, x) \phi_{i,a}(x, 0) = 0$ for any $y \in \mathcal{X} \setminus 0$. For $y, w \in \mathcal{X} \setminus 0$, it holds

$$\sum_{x \in \mathcal{X}} \hat{\phi}_{i,a}(y, x) \phi_{i,a}(x, w) = \sum_{x \in \mathcal{X}} b_i(x) \frac{\partial \log b_i(x)}{\partial \eta_{i,y}} \frac{\partial \log b_i(x)}{\partial \theta_{i,w}}$$

$$= \sum_{x \in \mathcal{X}} \frac{\partial b_i(x)}{\partial \eta_{i,y}} [t_{i,w}(x) - \eta_{i,w}]$$

$$= \frac{\partial \eta_{i,w}}{\partial \eta_{i,y}} - \eta_{i,w} \sum_{x \in \mathcal{X}} \frac{\partial b_i(x)}{\partial \eta_{i,y}} = \delta(y, w).$$

Then, one obtains the following theorem.

Theorem 7 (Loop calculus for non-binary alphabet). *For any stationary point $(\{b_i\}, \{b_a\})$ of the Bethe free energy, (10) holds where*

$$\mathcal{Z}(E') := \sum_{y \in (\mathcal{X} \setminus \{0\})^{|E'|} i \in V} \prod_{i \in V} \left\langle \prod_{a \in \partial i, (i,a) \in E'} \frac{\partial \log b_i(X_i)}{\partial \eta_{i,y_{i,a}}} \right\rangle_{b_i}$$

$$\cdot \prod_{a \in F} \left\langle \prod_{i \in \partial a, (i,a) \in E'} \frac{\partial \log b_i(X_i)}{\partial \theta_{i,y_{i,a}}} \right\rangle_{b_a}.$$

This theorem is useful for understanding the approximations to be proposed in Sec. VI.

V. EDGE ZETA FUNCTION AND THE DETERMINANT OF HESSIAN OF THE BETHE FREE ENERGY

A. Edge zeta function

In this section, the edge zeta function and Watanabe-Fukumizu formula are introduced [8].

Definition 8. For $(i,a) \in E$ and $(j,b) \in E$, $(i \rightarrow a) \rightarrow (j \rightarrow b) \stackrel{\text{def}}{\iff} j \in \partial a, i \neq j, a \neq b$.

Definition 9 (Prime cycle). The closed walk $e_1 \rightarrow e_2 \rightarrow \dots \rightarrow e_n \rightarrow e_1$ is said to be a prime cycle if and only if it cannot be expressed as a power of another walk. Prime cycles are identified up to cyclic permutations.

Definition 10 (Edge zeta function [8]). Let $r_{i \rightarrow a}$ be a natural number associated with an edge $(i,a) \in E$ and $u_{(i \rightarrow a), (j \rightarrow b)}$ be an $r_{i \rightarrow a} \times r_{j \rightarrow b}$ matrix for $(i \rightarrow a) \rightarrow (j \rightarrow b)$. Then, the edge zeta function is defined as

$$\zeta(\mathbf{u}) := \prod_{\substack{p=(e_1 \rightarrow e_2 \rightarrow \dots \rightarrow e_n \rightarrow e_1) \\ \text{is a prime cycle}}} \frac{1}{\det(I_{r_{e_1}} - u_{e_1, e_2} u_{e_2, e_3} \dots u_{e_n, e_1})}$$

where I_r is the identity matrix of size r .

If a factor graph includes more than one cycle, the number of prime cycles is infinite. Hence, it is difficult to evaluate $\zeta(\mathbf{u})$ from the definition. The following lemma is generally useful for evaluating $\zeta(\mathbf{u})$.

Lemma 11 (Bass's formula). *It holds $\zeta(\mathbf{u}) = \det(\mathcal{I}_{|E|} - \mathcal{M}(\mathbf{u}))^{-1}$ where $\mathcal{I}_{|E|}$ is the identity matrix of size $|E|$ as a block matrix, and where $\mathcal{M}(\mathbf{u})_{e,e'} := u_{e,e'}$ if $e \rightarrow e'$ and $\mathcal{M}(\mathbf{u})_{e,e'} := 0$ otherwise.*

Furthermore, the other expression of $\zeta(\mathbf{u})$ is known on some condition.

Lemma 12 (Ihara-Bass formula [8]). *Let r_i be a natural number associated with a variable node $i \in V$. When $u_{(i \rightarrow a), (j \rightarrow b)}$ is an $r_i \times r_j$ matrix independent of b and denoted by $u_{i \rightarrow j}^a$,*

$$\zeta(\mathbf{u})^{-1} = \det(\mathcal{I}_N - \mathcal{D} + \mathcal{W}) \prod_{a \in F} \det(\mathcal{U}^a)$$

where \mathcal{D} is an $N \times N$ block diagonal matrix defined by $\mathcal{D}_{i,i} := d_i I_{r_i}$, where \mathcal{U}^a is a $d_a \times d_a$ block matrix defined by $\mathcal{U}_{i,i}^a := I_{r_i}$ and $\mathcal{U}_{i,j}^a := u_{i \rightarrow j}^a$ for $i \neq j$, and where \mathcal{W} is an $N \times N$ block matrix defined by $\mathcal{W}_{i,j} := \sum_{a: \{i,j\} \subseteq \partial a} w_{i \rightarrow j}^a$. Here, $w_{i \rightarrow j}^a := (\mathcal{U}^{a^{-1}})_{i,j}$.

B. Determinant of Hessian of the Bethe free energy

In this subsection, $\{b_i\}$ and $\{b_a\}$ are assumed to be members of an arbitrary fixed parametric family of distributions. The alphabet \mathcal{X} is not necessarily finite. For $i \in V$, b_i has a parameter η_i . For $a \in F$, b_a has a parameter $\eta_a = (\eta_{\langle a \rangle}, (\eta_i)_{i \in \partial a})$. The condition (1) is assumed to be satisfied for any coordinate $(\{\eta_i\}, \{\eta_{\langle a \rangle}\})$. In the following, a parameter η is denoted by the normal font η for the simplicity. For $i \in V$ and $a \in F$, let H_i and H_a be the Shannon entropy of b_i and b_a , respectively. The notation $B \succ 0$ means that a matrix B is positive definite.

Lemma 13. *For $(\{\eta_i\}, \{\eta_{\langle a \rangle}\})$ satisfying*

$$\frac{\partial H_i}{\partial \eta_i \partial \eta_i} \succ 0, \quad \forall i \in V \quad \frac{\partial H_a}{\partial \eta_{\langle a \rangle} \partial \eta_{\langle a \rangle}} \succ 0, \quad \forall a \in F$$

it holds that

$$\det(\nabla^2 F_{\text{Bethe}}(\{\eta_i\}, \{\eta_{\langle a \rangle}\})) = \prod_{i \in V} \det\left(\frac{\partial H_i}{\partial \eta_i \partial \eta_i}\right) \prod_{a \in F} \det\left(\frac{\partial H_a}{\partial \eta_{\langle a \rangle} \partial \eta_{\langle a \rangle}}\right) \det(\mathcal{I}_N - \mathcal{D} + \mathcal{G})$$

where

$$\mathcal{G}_{i,j} := \left(\frac{\partial H_i}{\partial \eta_i \partial \eta_i}\right)^{-\frac{1}{2}} \left[\sum_{a \in \partial i \cap \partial j} \left(\frac{\partial H_a}{\partial \eta_i \partial \eta_j}\right) - \frac{\partial H_a}{\partial \eta_i \partial \eta_{\langle a \rangle}} \left(\frac{\partial H_a}{\partial \eta_{\langle a \rangle} \partial \eta_{\langle a \rangle}}\right)^{-1} \frac{\partial H_a}{\partial \eta_{\langle a \rangle} \partial \eta_j} \right] \left(\frac{\partial H_j}{\partial \eta_j \partial \eta_j}\right)^{-\frac{1}{2}}.$$

Proof. It is easy to see that

$$\begin{aligned} \frac{\partial F_{\text{Bethe}}}{\partial \eta_i \partial \eta_j} &= \sum_{a \in \partial i \cap \partial j} \frac{\partial H_a}{\partial \eta_i \partial \eta_j} - \delta_{i,j} (d_i - 1) \frac{\partial H_i}{\partial \eta_i \partial \eta_i} \\ \frac{\partial F_{\text{Bethe}}}{\partial \eta_{\langle a \rangle} \partial \eta_{\langle b \rangle}} &= \delta_{a,b} \frac{\partial H_a}{\partial \eta_{\langle a \rangle} \partial \eta_{\langle b \rangle}}, \quad \frac{\partial F_{\text{Bethe}}}{\partial \eta_i \partial \eta_{\langle a \rangle}} = \frac{\partial H_a}{\partial \eta_i \partial \eta_{\langle a \rangle}}. \end{aligned}$$

Let \mathcal{V} be a block diagonal matrix defined by

$$\mathcal{V}_{i,i} := \frac{\partial H_i}{\partial \eta_i \partial \eta_i}, \quad \mathcal{V}_{a,a} := \frac{\partial H_a}{\partial \eta_{\langle a \rangle} \partial \eta_{\langle a \rangle}}$$

and $\mathcal{C} := \nabla^2 F_{\text{Bethe}}(\{\eta_i\}, \{\eta_{\langle a \rangle}\}) - \mathcal{V}$. Then, one obtains

$$\nabla^2 F_{\text{Bethe}}(\{\eta_i\}, \{\eta_{\langle a \rangle}\}) = \mathcal{V}^{\frac{1}{2}} (\mathcal{I}_{N+|F|} + \mathcal{V}^{-\frac{1}{2}} \mathcal{C} \mathcal{V}^{-\frac{1}{2}}) \mathcal{V}^{\frac{1}{2}}.$$

For $\mathcal{F} := \mathcal{V}^{-\frac{1}{2}} \mathcal{C} \mathcal{V}^{-\frac{1}{2}}$, it holds that

$$\begin{aligned} \mathcal{F}_{i,j} &= \mathcal{V}_{i,i}^{-\frac{1}{2}} \sum_{a \in \partial i \cap \partial j} \frac{\partial H_a}{\partial \eta_i \partial \eta_j} \mathcal{V}_{j,j}^{-\frac{1}{2}} - \delta_{i,j} d_i I_{r_i}, \quad \mathcal{F}_{a,b} = 0 \\ \mathcal{F}_{i,a} &= \mathcal{V}_{i,i}^{-\frac{1}{2}} \frac{\partial H_a}{\partial \eta_i \partial \eta_{\langle a \rangle}} \mathcal{V}_{a,a}^{-\frac{1}{2}}, \quad \mathcal{F}_{a,i} = \mathcal{V}_{a,a}^{-\frac{1}{2}} \frac{\partial H_a}{\partial \eta_{\langle a \rangle} \partial \eta_i} \mathcal{V}_{i,i}^{-\frac{1}{2}}. \end{aligned}$$

From $\det(\nabla^2 F_{\text{Bethe}}(\{\eta_i\}, \{\eta_{\langle a \rangle}\})) = \det(\mathcal{V}) \det(\mathcal{I}_{N+|F|} + \mathcal{F})$ and

$$\det(\mathcal{V}) = \prod_{i \in V} \det\left(\frac{\partial H_i}{\partial \eta_i \partial \eta_i}\right) \prod_{a \in F} \det\left(\frac{\partial H_a}{\partial \eta_{\langle a \rangle} \partial \eta_{\langle a \rangle}}\right)$$

we only have to prove $\det(\mathcal{I}_{N+|F|} + \mathcal{F}) = \det(\mathcal{I}_N - \mathcal{D} + \mathcal{G})$. For $u \times u$, $u \times v$ and $v \times u$ matrices A , B and C , respectively, it holds $\begin{bmatrix} A & B \\ C & I_v \end{bmatrix} \begin{bmatrix} I_u & 0 \\ -C & I_v \end{bmatrix} = \begin{bmatrix} A - BC & B \\ 0 & I_v \end{bmatrix}$ and hence

$$\det\left(\begin{bmatrix} A & B \\ C & I_v \end{bmatrix}\right) = \det(A - BC). \text{ Therefore,}$$

$$\begin{aligned} \det(\mathcal{I}_{N+|F|} + \mathcal{F}) &= \det(\mathcal{I}_N + \mathcal{F}_{\text{VV}} - \mathcal{F}_{\text{VF}} \mathcal{F}_{\text{VF}}^t) \\ &= \det(\mathcal{I}_N - \mathcal{D} + \mathcal{G}). \quad \square \end{aligned}$$

For an exponential family, the determinant of Hessian of the Bethe free energy is connected to the edge zeta function.

Lemma 14 (Watanabe-Fukumizu formula [8]). *Let $\{\eta_i\}$ and $\{\eta_a\}$ be the expectation parameters for $\{b_i\}$ and $\{b_a\}$, respectively. Let $\{t_i\}$ and $\{t_a\}$ be the sufficient statistics for $\{b_i\}$ and $\{b_a\}$, respectively. Then, it holds*

$$\begin{aligned} \zeta(\mathbf{u})^{-1} &= \det(\nabla^2 F_{\text{Bethe}}(\{\eta_i\}, \{\eta_{\langle a \rangle}\})) \\ &\quad \cdot \prod_{i \in V} \det(\text{Var}_{b_i}[t_i(X_i)])^{1-d_i} \prod_{a \in F} \det(\text{Var}_{b_a}[t_a(X_{\partial a})])^{-1} \end{aligned}$$

where r_i is the number of parameters of b_i for $i \in V$ and where

$$\begin{aligned} u_{i \rightarrow j}^a &= \text{Cor}_{b_a}[t_i(X_i), t_j(X_j)] \\ &:= \text{Var}_{b_i}[t_i(X_i)]^{-\frac{1}{2}} \text{Cov}_{b_a}[t_i(X_i), t_j(X_j)] \text{Var}_{b_j}[t_j(X_j)]^{-\frac{1}{2}}. \end{aligned}$$

Here, $\text{Cov}_{b_a}[t_i(X_i), t_j(X_j)]$ is a matrix whose (k, l) -element is

$$\mathbb{E}[(t_{i,k}(X_i) - \mathbb{E}[t_{i,k}(X_i)])(t_{j,l}(X_j) - \mathbb{E}[t_{j,l}(X_j)])].$$

VI. THE NEW APPROXIMATIONS VIA ASYMPTOTIC EXPANSION

A. The series of approximations

Recently, a new characterization of the Bethe free energy is shown by Vontobel [2]. Let $Z(M)$ be the random variable corresponding to the partition function of a uniformly chosen random graph cover of the original factor graph where M is the number of copies of the original graph in the graph cover. See [2] for a detailed definition of a graph cover. Then, the Bethe free energy appears naturally as follows.

Lemma 15 ([2]). It holds $\mathbb{E}[Z(M)] = Z_{\text{Bethe}}^{M+o(M)}$ as $M \rightarrow \infty$.

This result gives a new characterization of the Bethe free energy. More detailed analysis is obtained as follows.

Lemma 16 ([9]). Let $\mathcal{S}(f)$ be the support of a function f . Let \mathcal{B} be the set of minima of $F_{\text{Bethe}}(\{b_i\}, \{b_a\})$. Assume $|\mathcal{B}| < \infty$ and

$$b_i^*(x) > 0, \forall x \in \mathcal{X}, \forall i \in V, \quad b_a^*(\mathbf{x}) > 0, \forall \mathbf{x} \in \mathcal{S}(f_a), \forall a \in F$$

$$\det(\nabla^2 F_{\text{Bethe}}(\{\eta_i^*\}, \{\eta_a^*\})) > 0$$

for all $(\{b_i^*\}, \{b_a^*\}) \in \mathcal{B}$. Then, it holds that as $M \rightarrow \infty$

$$\mathbb{E}[Z(M)] = Z_{\text{Bethe}}^M \sum_{(\{b_i^*\}, \{b_a^*\}) \in \mathcal{B}} \left(\det(\nabla^2 F_{\text{Bethe}}(\{\eta_i^*\}, \{\eta_a^*\})) \right)^{-\frac{1}{2}} (1 + o(1))$$

$$\cdot \prod_{i \in V} \prod_{x \in \mathcal{X}} b_i^*(x)^{1-d_i} \prod_{a \in F} \prod_{\mathbf{x} \in \mathcal{S}(f_a)} b_a^*(\mathbf{x}) \Bigg)$$

$$= Z_{\text{Bethe}}^M \sum_{(\{b_i^*\}, \{b_a^*\}) \in \mathcal{B}} \sqrt{\zeta(\mathbf{u})} (1 + o(1))$$

where $u_{i \rightarrow j}^a = \text{Cor}_{b_a^*}[t_i(X_i), t_j(X_j)]$ and where $\eta_i = \{b_i(x)\}_{x \in \mathcal{X} \setminus 0}$ and $\eta_a = T_a(\{b_a(\mathbf{x})\}_{\mathbf{x} \in \mathcal{S}(f_a)})$. Here, for $a \in F$, $T_a(\eta)$ denotes an any subset of η with which $\{\eta_i\}_{i \in \partial a}$ can be regarded as an expectation parameter for b_a .

Furthermore, except for cases including singularity, the expectation of a partition function of a graph cover has the asymptotic expansion

$$\log \mathbb{E}[Z(M)] \sim M \log Z_{\text{Bethe}} + \sum_{k=0}^{\infty} \frac{g_k}{\sqrt{M^k}}$$

where $g_0 := \log \sum_{(\{b_i^*\}, \{b_a^*\}) \in \mathcal{B}} \sqrt{\zeta(\mathbf{u})}$ and where $\{g_k\}_{k=1,2,\dots}$, are some constants similarly to the Edgeworth expansion [10]. Based on the above divergent series, we demonstrate the following series of approximations by letting $M = 1$, which may seem shameful

$$\begin{aligned} \log Z &\approx \log Z_{\text{Bethe}} \\ \log Z &\approx \log Z_{\text{Bethe}} + g_0 \\ \log Z &\approx \log Z_{\text{Bethe}} + g_0 + g_1 \\ &\vdots \end{aligned}$$

Definition 17 (Asymptotic Bethe approximation). For $m = 0, 1, \dots$, the asymptotic Bethe approximation of order m is defined as

$$Z_{\text{AB}}^{(m)} := Z_{\text{Bethe}} \exp \left\{ \sum_{k=0}^{m-1} g_k \right\}.$$

B. Asymptotic exactness of the asymptotic Bethe approximation of order 1

In this subsection, we show cases in which the asymptotic Bethe approximation $Z_{\text{AB}}^{(1)}$ is asymptotically better in some limit than the Bethe approximation Z_{Bethe} . Let the set L_2 of loops be

$$L_2 := \{E' \subseteq E \mid E' \text{ is connected, } d_o(E') = 0 \text{ or } 2, \\ \forall o \in V \cup F\}.$$

The set L_2 of loops is a subset of the set G of generalized loops. For the binary case, a contribution $\mathcal{Z}(E')$ in Lemma 6 of a loop $E' \in L_2$ is

$$\mathcal{Z}(E') = \text{Cor}_{b_{a_\ell}}[X_{i_\ell}, X_{i_1}] \prod_{k=1}^{\ell-1} \text{Cor}_{b_{a_k}}[X_{i_k}, X_{i_{k+1}}]$$

where $(i_\ell \rightarrow a_\ell) \rightarrow (i_1 \rightarrow a_1) \rightarrow (i_2 \rightarrow a_2) \rightarrow \dots \rightarrow (i_\ell \rightarrow a_\ell)$ forms the loop E' and where $\ell := |E'|$. In the above, weight $\text{Cor}_{b_a}[X_i, X_j]$ of the connection of edges $(i_k \rightarrow a_k) \rightarrow (i_{k+1} \rightarrow a_{k+1})$ is the same as $u_{i_k \rightarrow j_{k+1}}^{a_k}$ in the edge zeta function used for $Z_{\text{AB}}^{(1)}$ for $k = 1, \dots, \ell$.

For non-binary cases, the partition function of a single-cycle graph can also be calculated easily by using the well-known method of the transfer matrix.

Lemma 18 (Transfer matrices).

$$\sum_{\mathbf{x} \in \mathcal{X}^N} f_N(x_N, x_1) \prod_{i=1}^{N-1} f_i(x_i, x_{i+1}) = \text{tr}(F^{(1)} F^{(2)} \dots F^{(N)})$$

where $F^{(i)}$ is a $|\mathcal{X}| \times |\mathcal{X}|$ matrix with $F_{x,x'}^{(i)} = f_i(x, x')$.

From this lemma and Lemma 5, one obtains $\mathcal{Z}(E')$ in Theorem 7 for $E' \in L_2$ as

$$\begin{aligned} \mathcal{Z}(E') &= \text{tr}(\text{Cor}_{b_{a_1}}[t_{i_1}(X_{i_1}), t_{i_2}(X_{i_2})] \text{Cor}_{b_{a_2}}[t_{i_2}(X_{i_2}), t_{i_3}(X_{i_3})] \\ &\quad \dots \text{Cor}_{b_{a_\ell}}[t_{i_\ell}(X_{i_\ell}), t_{i_1}(X_{i_1})]). \end{aligned} \quad (11)$$

First, we consider the simplest non-trivial example, namely, a single-cycle factor graph.

Example 19 (Single-cycle factor graph). For a single-cycle graph, the Bethe free energy is convex with respect to the expectation parameters and hence the stationary point is unique [8]. For the unique solution $\{b_i^*\}, \{b_a^*\}$, one obtains from Theorem 7 and (11) that

$$\frac{Z}{Z_{\text{Bethe}}(\{b_i^*\}, \{b_a^*\})} = 1 + \text{tr}(A)$$

where

$$\begin{aligned} A &:= \text{Cor}_{b_{a_1}^*}[t_1(X_1), t_2(X_2)] \text{Cor}_{b_{a_2}^*}[t_2(X_2), t_3(X_3)] \\ &\quad \dots \text{Cor}_{b_{a_N}^*}[t_N(X_N), t_1(X_1)]. \end{aligned}$$

On the other hand, the square root of the edge zeta function is

$$\sqrt{\zeta(\mathbf{u})} = \frac{1}{\det(I_{|\mathcal{X}|-1} - A)}.$$

From $\det(I_{|\mathcal{X}|-1} - A) = 1 - \text{tr}(A) + O(\rho(A)^2)$ as $A \rightarrow 0$, where $\rho(A)$ denotes the spectral radius of A , one obtains the following asymptotic equality

$$\begin{aligned} \sqrt{\zeta(\mathbf{u})} &= \frac{1}{1 - \text{tr}(A) + O(\rho(A)^2)} \\ &= 1 + \text{tr}(A) + O(\rho(A)^2) = \frac{Z}{Z_{\text{Bethe}}} + O(\rho(A)^2). \end{aligned}$$

Hence, $\sqrt{\zeta(\mathbf{u})}$ is an accurate approximation for Z/Z_{Bethe} whenever the matrix A is close to zero. \square

For a factor graph including multiple cycles, we consider the case that the functions $\{f_a\}$ have a parameter $\beta > 0$. We assume that there is a unique minimum of the Bethe free energy. Furthermore, we assume that $Z/Z_{\text{Bethe}} \rightarrow 1$ as $\beta \rightarrow 0$, and that the contribution of loops dominates Z/Z_{Bethe} , i.e.,

$$\frac{Z}{Z_{\text{Bethe}}} = 1 + \sum_{E' \in L_2 \setminus \{\emptyset\}} \mathcal{Z}(E') + o\left(\sum_{E' \in L_2 \setminus \{\emptyset\}} \mathcal{Z}(E')\right) \quad (12)$$

as $\beta \rightarrow 0$. On the other hand, one obtains

$$\sqrt{\zeta(\mathbf{u})} = 1 + \sum_{E' \in L_2 \setminus \{\emptyset\}} \mathcal{Z}(E') + o\left(\sum_{E' \in L_2 \setminus \{\emptyset\}} \mathcal{Z}(E')\right).$$

Hence, the square root of the edge zeta function $\sqrt{\zeta(\mathbf{u})}$ gives the dominant terms in Z/Z_{Bethe} in the limit $\beta \rightarrow 0$ for an arbitrary topological graph. The derivation of the above equation is omitted due to the lack of the space.

Example 20 (High-temperature expansion for the Ising model). For $J_{i,j} \in \mathbb{R}$ and $\beta > 0$, the partition function of the Ising model is

$$Z = \sum_{\mathbf{x} \in \{+1, -1\}^V} \exp\left\{\beta \left(\sum_{(i,j) \in F} J_{i,j} x_i x_j + \sum_{i \in V} h_i x_i \right)\right\}. \quad (13)$$

When $h_i = 0$ for all $i \in V$, the set of uniform messages is the trivial solution of the BP equations (3). Although it is not the unique solution of (3), if β is sufficiently small, the set of uniform messages is the unique minimum of the Bethe free energy. The Fourier transform is equivalent to the Bethe transform for the uniform messages. From Lemma 6, or equivalently from the MacWilliams identity, one obtains

$$Z = 2^N \prod_{(i,j) \in F} \cosh(\beta J_{i,j}) \sum_{\substack{E' \subseteq E, \\ d_i(E') \text{ is even}}} \prod_{(i,j) \in E'} \tanh(\beta J_{i,j}).$$

In this case, the Bethe free energy evaluated with the uniform messages is $2^N \prod_{(i,j) \in F} \cosh(\beta J_{i,j})$. Since (12) is satisfied in this case, $Z_{\text{AB}}^{(1)}$ is a more accurate approximation than Z_{Bethe} in the limit $\beta \rightarrow 0$. When $h_i \neq 0$, the uniform messages are not a solution of the BP equations (3). In that case, from (2), $\text{Cor}_{b_a}[X_i, X_j]$ is equal to

$$\frac{\sinh(2\beta J_{i,j})}{\sqrt{\cosh(2l_i) + \cosh(2\beta J_{i,j})} \sqrt{\cosh(2l_j) + \cosh(2\beta J_{i,j})}}$$

where $m_{i \rightarrow a} \propto \exp\{l_i\}$ and $m_{j \rightarrow a} \propto \exp\{l_j\}$ [8]. Since $|\text{Cor}_{b_a}[X_i, X_j]|$ takes the maximum $|\tanh(\beta J_{i,j})|$ at $l_i = l_j = 0$, as $\beta \rightarrow 0$, $\text{Cor}_{b_a}[X_i, X_j] \rightarrow 0$ and hence, (12) is satisfied. On the other hand, as $h_i \rightarrow \infty$, it also holds $\text{Cor}_{b_a}[X_i, X_j] \rightarrow 0$ since $l_i \rightarrow \pm\infty$ as $h_i \rightarrow \pm\infty$. Hence, the condition (12) is also satisfied in the limit $h_i \rightarrow \pm\infty$ for all $i \in V$. \square

Remark 21. Let us consider the (not bipartite) graph in which a degree-two factor node in the Ising model (13) is replaced by an edge. Due to Kac and Ward [11], if the graph is planar and if $h_i = 0$ for all $i \in V$, the partition function Z can be calculated in polynomial time from the

beautiful equation $Z = Z_{\text{Bethe}} / \sqrt{\zeta(\mathbf{u})}$ where $u_{(i \rightarrow a), (j \rightarrow b)} = \tanh(\beta J_{i,j}) \exp\{\sqrt{-1} \gamma_{(i \rightarrow a), (j \rightarrow b)} / 2\}$. Here, $\gamma_{(i \rightarrow a), (j \rightarrow b)}$ denotes the angle of the edge connection $(i \rightarrow a) \rightarrow (j \rightarrow b)$. This equation exists only when the graph is a certain type of planar graph and the magnetic field h_i is zero for all $i \in V$. \square

Even if each $\text{Cor}_{b_a}[X_i, X_j]$ does not go to 0, if the product of them along a loop goes to 0, the condition (12) can be satisfied. Some sparse factor graphs can satisfy this condition in the large-size limit $N \rightarrow \infty$. The justification of the approximation $Z_{\text{AB}}^{(1)}$ for some sparse factor graphs is an open problem.

When the Hessian of the Bethe free energy is not positive definite at some critical temperature β_c , the edge zeta function diverges. This situation is considered as a finite-size analogue of the second-order phase transition. Similarly, if the minima of the Bethe free energy discontinuously jumps at β_c , $Z_{\text{AB}}^{(m)}$ is discontinuous at β_c for $m \geq 1$. This situation is considered as a finite-size analogue of the first-order phase transition. In these cases, it is better to consider another limit for β around β_c , e.g., $\delta = (\beta - \beta_c)/N^c$ is fixed for some $c > 0$ [12].

For the three-body Ising model without magnetic field, $\sqrt{\zeta(\mathbf{u})} = 1$ for the uniform messages and hence $Z_{\text{AB}}^{(1)} = Z_{\text{Bethe}}$. We can intuitively guess that $Z_{\text{AB}}^{(2)}$ takes account of contributions of the set of connected generalized loops in which all degrees must be at most three since the derivation of g_1 requires derivation of third-order derivatives of the Bethe free energy, which includes the third-order statistics in Theorem 7. This idea can be generalized to $Z_{\text{AB}}^{(m)}$ for any $m \geq 1$.

BIBLIOGRAPHY

- [1] M. Mezard and A. Montanari, *Information, Physics, and Computation*. Oxford University Press, USA, 2009.
- [2] P. Vontobel, "Counting in graph covers: A combinatorial characterization of the Bethe entropy function," 2010. [Online]. Available: <http://arxiv.org/abs/1012.0065v1>
- [3] J. Yedidia, W. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2282–2312, 2005.
- [4] V. Chernyak and M. Chertkov, "Loop calculus and belief propagation for q-ary alphabet: Loop tower," in *Proc. 2007 IEEE Int. Symposium on Inform. Theory, Nice, France*, Jun. 24–29, 2007, pp. 316–320.
- [5] G. Forney and P. Vontobel, "Partition functions of normal factor graphs," in *Proc. 2011 IEEE Inf. Theory and App. Workshop, La Jolla, CA*, 2011.
- [6] M. Chertkov and V. Chernyak, "Loop calculus in statistical physics and information science," *Physical Review E*, vol. 73, no. 6, p. 065102, 2006.
- [7] E. Sudderth, M. Wainwright, and A. Willsky, "Loop series and Bethe variational bounds in attractive graphical models," in *Advances in neural information processing systems*, vol. 20, 2008, pp. 1425–1432.
- [8] Y. Watanabe, "Discrete geometric analysis of message passing algorithm on graphs," Ph.D. dissertation, The Graduate University for Advanced Studies, Mar. 2010.
- [9] R. Mori and T. Tanaka, "Central approximation in statistical physics and information theory," in *Proc. 2012 IEEE Int. Symposium on Inform. Theory, St. Petersburg, Russia*, Jul. 1–6, 2012, pp. 1652–1656.
- [10] J. Chambers, "On methods of asymptotic approximation for multivariate distributions," *Biometrika*, vol. 54, no. 3-4, pp. 367–383, Dec. 1967.
- [11] M. Kac and J. Ward, "A combinatorial solution of the two-dimensional Ising model," *Physical Review*, vol. 88, pp. 1332–1337, Dec. 1952.
- [12] G. Parisi, F. Ritort, and F. Slanina, "Critical finite-size corrections for the Sherrington-Kirkpatrick spin glass," *Journal of Physics A: Mathematical and General*, vol. 26, no. 2, pp. 247–259, Jan. 1993.